Amazon

Exam Questions AWS-Certified-Machine-Learning-Specialty

AWS Certified Machine Learning - Specialty

NEW QUESTION 1
An e commerce company wants to launch a new cloud-based product recommendation feature for its web application. Due to data localization regulations, any sensitive data must not leave its on-premises data center, and the product recommendation model must be trained and tested using nonsensitive data only. Data transfer to the cloud must use IPsec. The web application is hosted on premises with a PostgreSQL database that contains all the data. The company wants the data to be uploaded securely to Amazon S3 each day for model retraining.
How should a machine learning specialist meet these requirements?

A. Create an AWS Glue job to connect to the PostgreSQL DB instanc
B. Ingest tables without sensitive data through an AWS Site-to-Site VPN connection directly into Amazon S3.
C. Create an AWS Glue job to connect to the PostgreSQL DB instanc
D. Ingest all data through an AWS Site- to-Site VPN connection into Amazon S3 while removing sensitive data using a PySpark job.
E. Use AWS Database Migration Service (AWS DMS) with table mapping to select PostgreSQL tables with no sensitive data through an SSL connectio
F. Replicate data directly into Amazon S3.
G. Use PostgreSQL logical replication to replicate all data to PostgreSQL in Amazon EC2 through AWS Direct Connect with a VPN connectio
H. Use AWS Glue to move data from Amazon EC2 to Amazon S3.

Answer: C

NEW QUESTION 2
A retail company wants to update its customer support system. The company wants to implement automatic routing of customer claims to different queues to prioritize the claims by category.
Currently, an operator manually performs the category assignment and routing. After the operator classifies and routes the claim, the company stores the claim's record in a central database. The claim's record includes the claim's category.
The company has no data science team or experience in the field of machine learning (ML). The company's small development team needs a solution that requires no ML expertise.
Which solution meets these requirements?

A. Export the database to a .csv file with two columns: claim_label and claim_tex
B. Use the Amazon SageMaker Object2Vec algorithm and the .csv file to train a mode
C. Use SageMaker to deploy the model to an inference endpoin
D. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.
E. Export the database to a .csv file with one column: claim_tex
F. Use the Amazon SageMaker Latent Dirichlet Allocation (LDA) algorithm and the .csv file to train a mode
G. Use the LDA algorithm to detect labels automaticall
H. Use SageMaker to deploy the model to an inference endpoin
I. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.
J. Use Amazon Textract to process the database and automatically detect two columns: claim_label and claim_tex
K. Use Amazon Comprehend custom classification and the extracted information to train the custom classifie
L. Develop a service in the application to use the Amazon Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.
M. Export the database to a .csv file with two columns: claim_label and claim_tex
N. Use Amazon Comprehend custom classification and the .csv file to train the custom classifie
O. Develop a service in the application to use the Amazon Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.

Answer: C

NEW QUESTION 3
A company ingests machine learning (ML) data from web advertising clicks into an Amazon S3 data lake. Click data is added to an Amazon Kinesis data stream by using the Kinesis Producer Library (KPL). The data is loaded into the S3 data lake from the data stream by using an Amazon Kinesis Data Firehose delivery stream. As the data volume increases, an ML specialist notices that the rate of data ingested into Amazon S3 is relatively constant. There also is an increasing backlog of data for Kinesis Data Streams and Kinesis Data Firehose to ingest.
Which next step is MOST likely to improve the data ingestion rate into Amazon S3?

A. Increase the number of S3 prefixes for the delivery stream to write to.
B. Decrease the retention period for the data stream.
C. Increase the number of shards for the data stream.
D. Add more consumers using the Kinesis Client Library (KCL).

Answer: C

NEW QUESTION 4
A medical imaging company wants to train a computer vision model to detect areas of concern on patients' CT scans. The company has a large collection of unlabeled CT scans that are linked to each patient and stored in an Amazon S3 bucket. The scans must be accessible to authorized users only. A machine learning engineer needs to build a labeling pipeline.
Which set of steps should the engineer take to build the labeling pipeline with the LEAST effort?

A. Create a workforce with AWS Identity and Access Management (IAM). Build a labeling tool on Amazon EC2 Queue images for labeling by using Amazon Simple Queue Service (Amazon SQS). Write the labeling instructions.
B. Create an Amazon Mechanical Turk workforce and manifest fil
C. Create a labeling job by using the built-in image classification task type in Amazon SageMaker Ground Trut
D. Write the labeling instructions.
E. Create a private workforce and manifest fil
F. Create a labeling job by using the built-in bounding box task type in Amazon SageMaker Ground Trut
G. Write the labeling instructions.
H. Create a workforce with Amazon Cognit
I. Build a labeling web application with AWS Amplif
J. Build a labeling workflow backend using AWS Lambd

K. Write the labeling instructions.

Answer: C

Explanation:
https://docs.aws.amazon.com/sagemaker/latest/dg/sms-workforce-private.html

NEW QUESTION 5
A web-based company wants to improve its conversion rate on its landing page Using a large historical dataset of customer visits, the company has repeatedly trained a multi-class deep learning network algorithm on Amazon SageMaker However there is an overfitting problem training data shows 90% accuracy in predictions, while test data shows 70% accuracy only
The company needs to boost the generalization of its model before deploying it into production to maximize conversions of visits to purchases
Which action is recommended to provide the HIGHEST accuracy model for the company's test and validation data?

A. Increase the randomization of training data in the mini-batches used in training.
B. Allocate a higher proportion of the overall data to the training dataset
C. Apply L1 or L2 regularization and dropouts to the training.
D. Reduce the number of layers and units (or neurons) from the deep learning network.

Answer: C

Explanation:
If this is a ComputerVision problem augmentation can help and we may consider A an option. However in analyzing customer historic data, there is no easy way to increase randomization in training. If you go deep into modelling and coding. When you build model with tensorflow/pytorch, most of the time the trainloader is already sampling in data in random manner (with shuffle enable). What we usually do to reduce overfitting is by adding dropout.
https://docs.aws.amazon.com/machine-learning/latest/dg/model-fit-underfitting-vs-overfitting.html

NEW QUESTION 6
A Machine Learning Specialist is implementing a full Bayesian network on a dataset that describes public transit in New York City. One of the random variables is discrete, and represents the number of minutes New Yorkers wait for a bus given that the buses cycle every 10 minutes, with a mean of 3 minutes.
Which prior probability distribution should the ML Specialist use for this variable?

A. Poisson distribution ,
B. Uniform distribution
C. Normal distribution
D. Binomial distribution

Answer: A

NEW QUESTION 7
A Machine Learning Specialist is creating a new natural language processing application that processes a dataset comprised of 1 million sentences The aim is to then run Word2Vec to generate embeddings of the sentences and enable different types of predictions
Here is an example from the dataset
"The quck BROWN FOX jumps over the lazy dog "
Which of the following are the operations the Specialist needs to perform to correctly sanitize and prepare the data in a repeatable manner? (Select THREE)

A. Perform part-of-speech tagging and keep the action verb and the nouns only
B. Normalize all words by making the sentence lowercase
C. Remove stop words using an English stopword dictionary.
D. Correct the typography on "quck" to "quick."
E. One-hot encode all words in the sentence
F. Tokenize the sentence into words.

Answer: BCF

NEW QUESTION 8
A large mobile network operating company is building a machine learning model to predict customers who are likely to unsubscribe from the service. The company plans to offer an incentive for these customers as the cost of churn is far greater than the cost of the incentive.
The model produces the following confusion matrix after evaluating on a test dataset of 100 customers: Based on the model evaluation results, why is this a viable model for production?

n = 100	PREDICTED CHURN Yes	PREDICTED CHURN No
ACTUAL Churn Yes	10	4
Actual No	10	76

A. The model is 86% accurate and the cost incurred by the company as a result of false negatives is less than the false positives.
B. The precision of the model is 86%, which is less than the accuracy of the model.
C. The model is 86% accurate and the cost incurred by the company as a result of false positives is less than the false negatives.
D. The precision of the model is 86%, which is greater than the accuracy of the model.

Answer: A

NEW QUESTION 9

A Machine Learning Specialist is using Apache Spark for pre-processing training data As part of the Spark pipeline, the Specialist wants to use Amazon SageMaker for training a model and hosting it Which of the following would the Specialist do to integrate the Spark application with SageMaker? (Select THREE)

A. Download the AWS SDK for the Spark environment
B. Install the SageMaker Spark library in the Spark environment.
C. Use the appropriate estimator from the SageMaker Spark Library to train a model.
D. Compress the training data into a ZIP file and upload it to a pre-defined Amazon S3 bucket.
E. Use the sageMakerMode
F. transform method to get inferences from the model hosted in SageMaker
G. Convert the DataFrame object to a CSV file, and use the CSV file as input for obtaining inferences from SageMaker.

Answer: DEF

NEW QUESTION 10
The Chief Editor for a product catalog wants the Research and Development team to build a machine learning system that can be used to detect whether or not individuals in a collection of images are wearing the company's retail brand The team has a set of training data
Which machine learning algorithm should the researchers use that BEST meets their requirements?

A. Latent Dirichlet Allocation (LDA)
B. Recurrent neural network (RNN)
C. K-means
D. Convolutional neural network (CNN)

Answer: C

NEW QUESTION 10
A manufacturer of car engines collects data from cars as they are being driven The data collected includes timestamp, engine temperature, rotations per minute (RPM), and other sensor readings The company wants to predict when an engine is going to have a problem so it can notify drivers in advance to get engine maintenance The engine data is loaded into a data lake for training
Which is the MOST suitable predictive model that can be deployed into production'?

A. Add labels over time to indicate which engine faults occur at what time in the future to turn this into a supervised learning problem Use a recurrent neural network (RNN) to train the model to recognize when an engine might need maintenance for a certain fault.
B. This data requires an unsupervised learning algorithm Use Amazon SageMaker k-means to cluster the data
C. Add labels over time to indicate which engine faults occur at what time in the future to turn this into a supervised learning problem Use a convolutional neural network (CNN) to train the model to recognize when an engine might need maintenance for a certain fault.
D. This data is already formulated as a time series Use Amazon SageMaker seq2seq to model the time series.
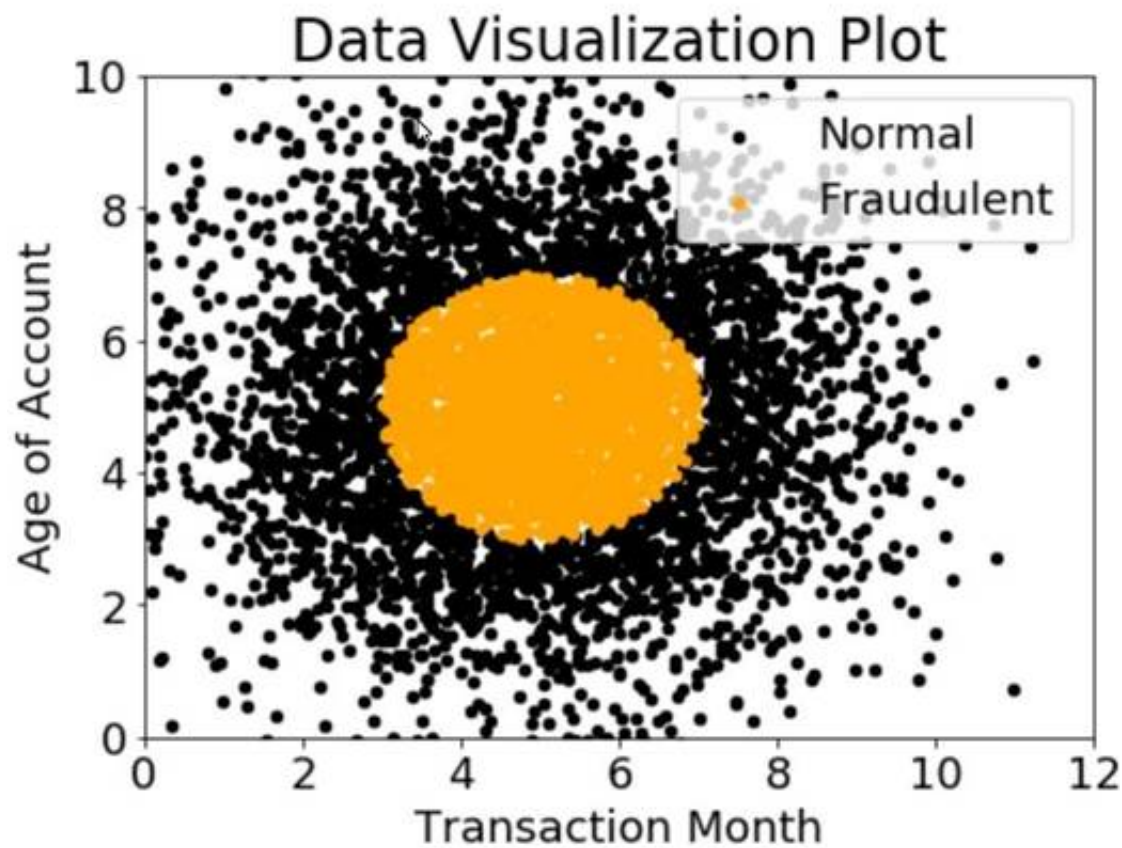
Answer: B

NEW QUESTION 11
An online reseller has a large, multi-column dataset with one column missing 30% of its data A Machine Learning Specialist believes that certain columns in the dataset could be used to reconstruct the missing data.
Which reconstruction approach should the Specialist use to preserve the integrity of the dataset?

A. Listwise deletion
B. Last observation carried forward
C. Multiple imputation
D. Mean substitution

Answer: C

NEW QUESTION 14
A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.

Based on this information which model would have the HIGHEST accuracy?

A. Long short-term memory (LSTM) model with scaled exponential linear unit (SELL))
B. Logistic regression
C. Support vector machine (SVM) with non-linear kernel
D. Single perceptron with tanh activation function

Answer: C

NEW QUESTION 17
A data science team is planning to build a natural language processing (NLP) application. The application's text preprocessing stage will include part-of-speech tagging and key phase extraction. The preprocessed text will be input to a custom classification algorithm that the data science team has already written and trained using Apache MXNet.
Which solution can the team build MOST quickly to meet these requirements?

A. Use Amazon Comprehend for the part-of-speech tagging, key phase extraction, and classification tasks.
B. Use an NLP library in Amazon SageMaker for the part-of-speech taggin
C. Use Amazon Comprehend for the key phase extractio
D. Use AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier.
E. Use Amazon Comprehend for the part-of-speech tagging and key phase extraction task
F. Use Amazon SageMaker built-in Latent Dirichlet Allocation (LDA) algorithm to build the custom classifier.
G. Use Amazon Comprehend for the part-of-speech tagging and key phase extraction task
H. Use AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier.

Answer: B

NEW QUESTION 20
A Data Scientist needs to migrate an existing on-premises ETL process to the cloud The current process runs at regular time intervals and uses PySpark to combine and format multiple large data sources into a single consolidated output for downstream processing
The Data Scientist has been given the following requirements for the cloud solution
* Combine multiple data sources
* Reuse existing PySpark logic
* Run the solution on the existing schedule
* Minimize the number of servers that will need to be managed
Which architecture should the Data Scientist use to build this solution?

A. Write the raw data to Amazon S3 Schedule an AWS Lambda function to submit a Spark step to a persistent Amazon EMR cluster based on the existing schedule Use the existing PySpark logic to run the ETL job on the EMR cluster Output the results to a "processed" location m Amazon S3 that is accessible tor downstream use
B. Write the raw data to Amazon S3 Create an AWS Glue ETL job to perform the ETL processing against the input data Write the ETL job in PySpark to leverage the existing logic Create a new AWS Glue trigger to trigger the ETL job based on the existing schedule Configure the output target of the ETL job to write to a "processed" location in Amazon S3 that is accessible for downstream use.
C. Write the raw data to Amazon S3 Schedule an AWS Lambda function to run on the existing schedule and process the input data from Amazon S3 Write the Lambda logic in Python and implement the existing PySpartc logic to perform the ETL process Have the Lambda function output the results to a "processed" location in Amazon S3 that is accessible for downstream use
D. Use Amazon Kinesis Data Analytics to stream the input data and perform realtime SQL queries against the stream to carry out the required transformations within the stream Deliver the output results to a "processed" location in Amazon S3 that is accessible for downstream use

Answer: A

NEW QUESTION 23
A Machine Learning Specialist needs to move and transform data in preparation for training Some of the data needs to be processed in near-real time and other data can be moved hourly There are existing Amazon EMR MapReduce jobs to clean and feature engineering to perform on the data
Which of the following services can feed data to the MapReduce jobs? (Select TWO)

A. AWSDMS
B. Amazon Kinesis
C. AWS Data Pipeline
D. Amazon Athena
E. Amazon ES

Answer: BC

Explanation:
https://aws.amazon.com/jp/emr/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-car

NEW QUESTION 28
A large company has developed a B1 application that generates reports and dashboards using data collected from various operational metrics The company wants to provide executives with an enhanced experience so they can use natural language to get data from the reports The company wants the executives to be able ask questions using written and spoken interlaces
Which combination of services can be used to build this conversational interface? (Select THREE)

A. Alexa for Business
B. Amazon Connect
C. Amazon Lex
D. Amazon Poly
E. Amazon Comprehend
F. Amazon Transcribe

Answer: BEF

NEW QUESTION 30
A Machine Learning Specialist is configuring automatic model tuning in Amazon SageMaker
When using the hyperparameter optimization feature, which of the following guidelines should be followed to improve optimization?
Choose the maximum number of hyperparameters supported by

A. Amazon SageMaker to search the largest number of combinations possible
B. Specify a very large hyperparameter range to allow Amazon SageMaker to cover every possible value.
C. Use log-scaled hyperparameters to allow the hyperparameter space to be searched as quickly as possible
D. Execute only one hyperparameter tuning job at a time and improve tuning through successive rounds of experiments

Answer: C

NEW QUESTION 35
A machine learning (ML) specialist wants to secure calls to the Amazon SageMaker Service API. The specialist has configured Amazon VPC with a VPC interface endpoint for the Amazon SageMaker Service API and is attempting to secure traffic from specific sets of instances and IAM users. The VPC is configured with a single public subnet.
Which combination of steps should the ML specialist take to secure the traffic? (Choose two.)

A. Add a VPC endpoint policy to allow access to the IAM users.
B. Modify the users' IAM policy to allow access to Amazon SageMaker Service API calls only.
C. Modify the security group on the endpoint network interface to restrict access to the instances.
D. Modify the ACL on the endpoint network interface to restrict access to the instances.
E. Add a SageMaker Runtime VPC endpoint interface to the VPC.

Answer: AC

NEW QUESTION 40
A Machine Learning Specialist is configuring Amazon SageMaker so multiple Data Scientists can access notebooks, train models, and deploy endpoints. To ensure the best operational performance, the Specialist needs to be able to track how often the Scientists are deploying models, GPU and CPU utilization on the deployed SageMaker endpoints, and all errors that are generated when an endpoint is invoked.
Which services are integrated with Amazon SageMaker to track this information? (Select TWO.)

A. AWS CloudTrail
B. AWS Health
C. AWS Trusted Advisor
D. Amazon CloudWatch
E. AWS Config

Answer: AD

NEW QUESTION 44
A Machine Learning Specialist is building a convolutional neural network (CNN) that will classify 10 types of animals. The Specialist has built a series of layers in a neural network that will take an input image of an animal, pass it through a series of convolutional and pooling layers, and then finally pass it through a dense and fully connected layer with 10 nodes The Specialist would like to get an output from the neural network that is a probability distribution of how likely it is that the input image belongs to each of the 10 classes
Which function will produce the desired output?

A. Dropout
B. Smooth L1 loss
C. Softmax
D. Rectified linear units (ReLU)

Answer: C

NEW QUESTION 48
A machine learning (ML) specialist must develop a classification model for a financial services company. A domain expert provides the dataset, which is tabular with 10,000 rows and 1,020 features. During exploratory data analysis, the specialist finds no missing values and a small percentage of duplicate rows. There are correlation scores of > 0.9 for 200 feature pairs. The mean value of each feature is similar to its 50th percentile.
Which feature engineering strategy should the ML specialist use with Amazon SageMaker?

A. Apply dimensionality reduction by using the principal component analysis (PCA) algorithm.
B. Drop the features with low correlation scores by using a Jupyter notebook.
C. Apply anomaly detection by using the Random Cut Forest (RCF) algorithm.
D. Concatenate the features with high correlation scores by using a Jupyter notebook.

Answer: C

NEW QUESTION 49
A real estate company wants to create a machine learning model for predicting housing prices based on a historical dataset. The dataset contains 32 features.
Which model will meet the business requirement?

A. Logistic regression
B. Linear regression
C. K-means
D. Principal component analysis (PCA)

Answer: B

NEW QUESTION 52
A large JSON dataset for a project has been uploaded to a private Amazon S3 bucket The Machine Learning Specialist wants to securely access and explore the data from an Amazon SageMaker notebook instance A new VPC was created and assigned to the Specialist
How can the privacy and integrity of the data stored in Amazon S3 be maintained while granting access to the Specialist for analysis?

A. Launch the SageMaker notebook instance within the VPC with SageMaker-provided internet access enabled Use an S3 ACL to open read privileges to the everyone group
B. Launch the SageMaker notebook instance within the VPC and create an S3 VPC endpoint for the notebook to access the data Copy the JSON dataset from Amazon S3 into the ML storage volume on the SageMaker notebook instance and work against the local dataset
C. Launch the SageMaker notebook instance within the VPC and create an S3 VPC endpoint for the notebook to access the data Define a custom S3 bucket policy to only allow requests from your VPC to access the S3 bucket
D. Launch the SageMaker notebook instance within the VPC with SageMaker-provided internet access enable
E. Generate an S3 pre-signed URL for access to data in the bucket

Answer: B

NEW QUESTION 56
A company that promotes healthy sleep patterns by providing cloud-connected devices currently hosts a sleep tracking application on AWS. The application collects device usage information from device users. The company's Data Science team is building a machine learning model to predict if and when a user will stop utilizing the company's devices. Predictions from this model are used by a downstream application that determines the best approach for contacting users.
The Data Science team is building multiple versions of the machine learning model to evaluate each version against the company's business goals. To measure long-term effectiveness, the team wants to run multiple versions of the model in parallel for long periods of time, with the ability to control the portion of inferences served by the models.
Which solution satisfies these requirements with MINIMAL effort?

A. Build and host multiple models in Amazon SageMake
B. Create multiple Amazon SageMaker endpoints, one for each mode
C. Programmatically control invoking different models for inference at the applicationlayer.
D. Build and host multiple models in Amazon SageMake
E. Create an Amazon SageMaker endpoint configuration with multiple production variant
F. Programmatically control the portion of the inferences served by the multiple models by updating the endpoint configuration.
G. Build and host multiple models in Amazon SageMaker Neo to take into account different types of medical device
H. Programmatically control which model is invoked for inference based on the medical device type.
I. Build and host multiple models in Amazon SageMake
J. Create a single endpoint that accesses multiple model
K. Use Amazon SageMaker batch transform to control invoking the different models through the single endpoint.

Answer: B

Explanation:
A/B testing with Amazon SageMaker is required in the Exam. In A/B testing, you test different variants of your models and compare how each variant performs. Amazon SageMaker enables you to test multiple models or model versions behind the `same endpoint` using `production variants`. Each production variant identifies a machine learning (ML) model and the resources deployed for hosting the model. To test multiple models by `distributing traffic` between them, specify the `percentage of the traffic` that gets routed to each model by specifying the `weight` for each `production variant` in the endpoint configuration.
https://docs.aws.amazon.com/sagemaker/latest/dg/model-ab-testing.html#model-testing-target-variant

NEW QUESTION 59
A Machine Learning Specialist is developing a daily ETL workflow containing multiple ETL jobs The workflow consists of the following processes
* Start the workflow as soon as data is uploaded to Amazon S3
* When all the datasets are available in Amazon S3, start an ETL job to join the uploaded datasets with multiple terabyte-sized datasets already stored in Amazon S3
* Store the results of joining datasets in Amazon S3

* If one of the jobs fails, send a notification to the Administrator Which configuration will meet these requirements?

A. Use AWS Lambda to trigger an AWS Step Functions workflow to wait for dataset uploads to complete in Amazon S3. Use AWS Glue to join the datasets Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure
B. Develop the ETL workflow using AWS Lambda to start an Amazon SageMaker notebook instance Use a lifecycle configuration script to join the datasets and persist the results in Amazon S3 Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure
C. Develop the ETL workflow using AWS Batch to trigger the start of ETL jobs when data is uploaded to Amazon S3 Use AWS Glue to join the datasets in Amazon S3 Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure
D. Use AWS Lambda to chain other Lambda functions to read and join the datasets in Amazon S3 as soon as the data is uploaded to Amazon S3 Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure

Answer: A

NEW QUESTION 63
A machine learning (ML) specialist needs to extract embedding vectors from a text series. The goal is to provide a ready-to-ingest feature space for a data scientist to develop downstream ML predictive models. The text consists of curated sentences in English. Many sentences use similar words but in different contexts. There are questions and answers among the sentences, and the embedding space must differentiate between them.
Which options can produce the required embedding vectors that capture word context and sequential QA information? (Choose two.)

A. Amazon SageMaker seq2seq algorithm
B. Amazon SageMaker BlazingText algorithm in Skip-gram mode
C. Amazon SageMaker Object2Vec algorithm
D. Amazon SageMaker BlazingText algorithm in continuous bag-of-words (CBOW) mode
E. Combination of the Amazon SageMaker BlazingText algorithm in Batch Skip-gram mode with a custom recurrent neural network (RNN)

Answer: AC

NEW QUESTION 67
A media company with a very large archive of unlabeled images, text, audio, and video footage wishes to index its assets to allow rapid identification of relevant content by the Research team. The company wants to use machine learning to accelerate the efforts of its in-house researchers who have limited machine learning expertise.
Which is the FASTEST route to index the assets?

A. Use Amazon Rekognition, Amazon Comprehend, and Amazon Transcribe to tag data into distinct categories/classes.
B. Create a set of Amazon Mechanical Turk Human Intelligence Tasks to label all footage.
C. Use Amazon Transcribe to convert speech to tex
D. Use the Amazon SageMaker Neural Topic Model (NTM) and Object Detection algorithms to tag data into distinct categories/classes.
E. Use the AWS Deep Learning AMI and Amazon EC2 GPU instances to create custom models for audio transcription and topic modeling, and use object detection to tag data into distinct categories/classes.

Answer: A

NEW QUESTION 69
Amazon Connect has recently been tolled out across a company as a contact call center The solution has been configured to store voice call recordings on Amazon S3
The content of the voice calls are being analyzed for the incidents being discussed by the call operators Amazon Transcribe is being used to convert the audio to text, and the output is stored on Amazon S3
Which approach will provide the information required for further analysis?

A. Use Amazon Comprehend with the transcribed files to build the key topics
B. Use Amazon Translate with the transcribed files to train and build a model for the key topics
C. Use the AWS Deep Learning AMI with Gluon Semantic Segmentation on the transcribed files to train and build a model for the key topics
D. Use the Amazon SageMaker k-Nearest-Neighbors (kNN) algorithm on the transcribed files to generate a word embeddings dictionary for the key topics

Answer: B

NEW QUESTION 72
A financial company is trying to detect credit card fraud. The company observed that, on average, 2% of credit card transactions were fraudulent. A data scientist trained a classifier on a year's worth of credit card transactions data. The model needs to identify the fraudulent transactions (positives) from the regular ones (negatives). The company's goal is to accurately capture as many positives as possible.
Which metrics should the data scientist use to optimize the model? (Choose two.)

A. Specificity
B. False positive rate
C. Accuracy
D. Area under the precision-recall curve
E. True positive rate

Answer: DE

NEW QUESTION 74
A company supplies wholesale clothing to thousands of retail stores. A data scientist must create a model that predicts the daily sales volume for each item for each store. The data scientist discovers that more than half of the stores have been in business for less than 6 months. Sales data is highly consistent from week to week. Daily data from the database has been aggregated weekly, and weeks with no sales are omitted from the current dataset. Five years (100 MB) of sales data is available in Amazon S3.
Which factors will adversely impact the performance of the forecast model to be developed, and which actions should the data scientist take to mitigate them? (Choose two.)

A. Detecting seasonality for the majority of stores will be an issu
B. Request categorical data to relate new stores with similar stores that have more historical data.
C. The sales data does not have enough varianc
D. Request external sales data from other industries to improve the model's ability to generalize.
E. Sales data is aggregated by wee
F. Request daily sales data from the source database to enable building a daily model.
G. The sales data is missing zero entries for item sale
H. Request that item sales data from the source database include zero entries to enable building the model.
I. Only 100 MB of sales data is available in Amazon S3. Request 10 years of sales data, which would provide 200 MB of training data for the model.

Answer: AB

NEW QUESTION 75
A Data Scientist is developing a machine learning model to classify whether a financial transaction is fraudulent. The labeled data available for training consists of 100,000 non-fraudulent observations and 1,000 fraudulent observations.
The Data Scientist applies the XGBoost algorithm to the data, resulting in the following confusion matrix when the trained model is applied to a previously unseen validation dataset. The accuracy of the model is 99.1%, but the Data Scientist needs to reduce the number of false negatives.

```
Predicted      0      1
Actual    0 99,966 | 34
          1    877 |123
```

Which combination of steps should the Data Scientist take to reduce the number of false negative predictions by the model? (Choose two.)

A. Change the XGBoost eval_metric parameter to optimize based on Root Mean Square Error (RMSE).
B. Increase the XGBoost scale_pos_weight parameter to adjust the balance of positive and negative weights.
C. Increase the XGBoost max_depth parameter because the model is currently underfitting the data.
D. Change the XGBoost eval_metric parameter to optimize based on Area Under the ROC Curve (AUC).
E. Decrease the XGBoost max_depth parameter because the model is currently overfitting the data.

Answer: BD

NEW QUESTION 80
A Data Scientist wants to gain real-time insights into a data stream of GZIP files. Which solution would allow the use of SQL to query the stream with the LEAST latency?

A. Amazon Kinesis Data Analytics with an AWS Lambda function to transform the data.
B. AWS Glue with a custom ETL script to transform the data.
C. An Amazon Kinesis Client Library to transform the data and save it to an Amazon ES cluster.
D. Amazon Kinesis Data Firehose to transform the data and put it into an Amazon S3 bucket.

Answer: A

NEW QUESTION 82
A machine learning specialist is developing a regression model to predict rental rates from rental listings. A variable named Wall_Color represents the most prominent exterior wall color of the property. The following is the sample data, excluding all other variables:

Property_ID	Wall_Color
1000	Red
1001	White
1002	Green

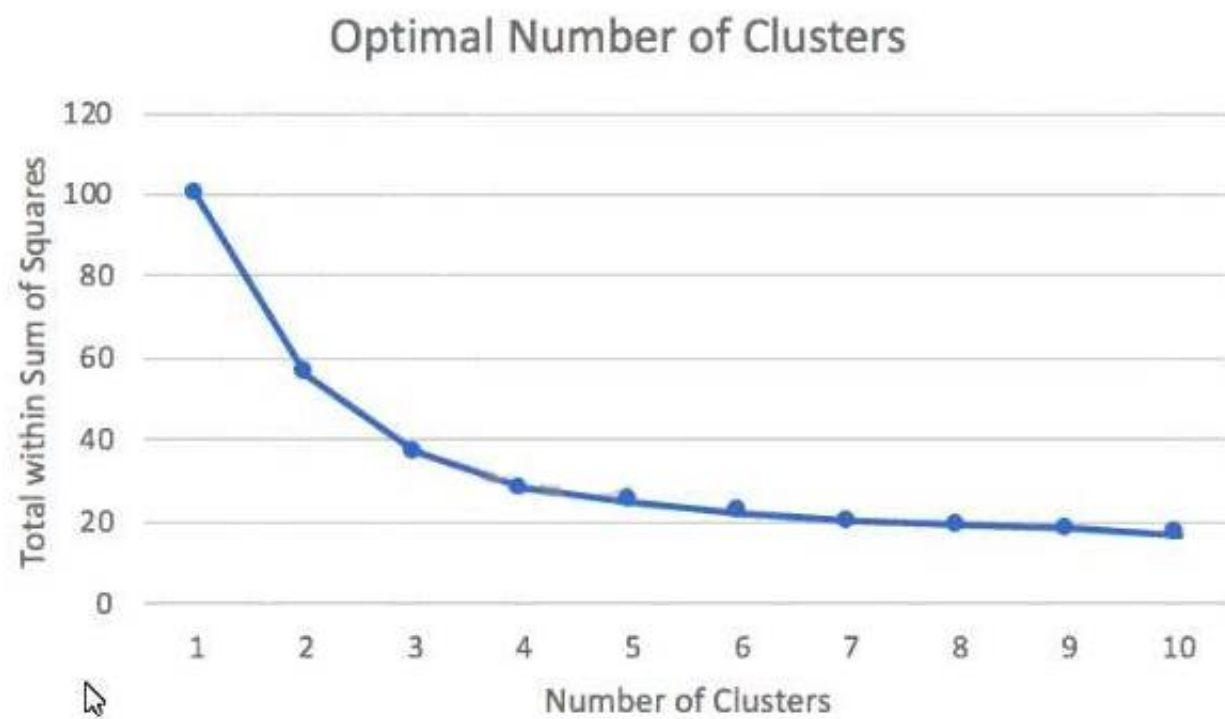The specialist chose a model that needs numerical input data.
Which feature engineering approaches should the specialist use to allow the regression model to learn from the Wall_Color data? (Choose two.)

A. Apply integer transformation and set Red = 1, White = 5, and Green = 10.
B. Add new columns that store one-hot representation of colors.
C. Replace the color name string by its length.
D. Create three columns to encode the color in RGB format.
E. Replace each color name by its training set frequency.

Answer: AD

NEW QUESTION 87
A Machine Learning Specialist prepared the following graph displaying the results of k-means for k = [1:10]

Optimal Number of Clusters

Considering the graph, what is a reasonable selection for the optimal choice of k?

A. 1
B. 4
C. 7
D. 10

Answer: C

NEW QUESTION 92
A Data Scientist is developing a binary classifier to predict whether a patient has a particular disease on a series of test results. The Data Scientist has data on 400 patients randomly selected from the population. The disease is seen in 3% of the population.
Which cross-validation strategy should the Data Scientist adopt?

A. A k-fold cross-validation strategy with k=5
B. A stratified k-fold cross-validation strategy with k=5
C. A k-fold cross-validation strategy with k=5 and 3 repeats
D. An 80/20 stratified split between training and validation

Answer: B

NEW QUESTION 96
A data scientist is training a text classification model by using the Amazon SageMaker built-in BlazingText algorithm. There are 5 classes in the dataset, with 300 samples for category A, 292 samples for category B, 240 samples for category C, 258 samples for category D, and 310 samples for category E.
The data scientist shuffles the data and splits off 10% for testing. After training the model, the data scientist generates confusion matrices for the training and test sets.

Training data confusion matrix

			Predicted class			
		A	B	C	D	E	Total
True class	A	270	0	0	0	0	270
	B	1	260	0	0	2	263
	C	0	0	111	100	5	216
	D	4	3	132	92	1	232
	E	0	0	2	3	274	279
	Total	275	263	245	195	282	1260

Test data confusion matrix

		Predicted class					
		A	B	C	D	E	Total
True class	A	9	1	0	0	0	10
	B	2	25	0	2	0	29
	C	10	2	11	10	1	34
	D	1	0	12	14	0	27
	E	9	1	4	1	25	40
	Total	31	29	27	27	26	140

What could the data scientist conclude form these results?

A. Classes C and D are too similar.
B. The dataset is too small for holdout cross-validation.
C. The data distribution is skewed.
D. The model is overfitting for classes B and E.

Answer: B

NEW QUESTION 98
A Machine Learning Specialist is attempting to build a linear regression model.
Given the displayed residual plot only, what is the MOST likely problem with the model?

A. Linear regression is inappropriat
B. The residuals do not have constant variance.
C. Linear regression is inappropriat
D. The underlying data has outliers.
E. Linear regression is appropriat
F. The residuals have a zero mean.
G. Linear regression is appropriat
H. The residuals have constant variance.

Answer: D

NEW QUESTION 102
A Data Science team within a large company uses Amazon SageMaker notebooks to access data stored in Amazon S3 buckets. The IT Security team is concerned that internet-enabled notebook instances create a security vulnerability where malicious code running on the instances could compromise data privacy. The company mandates that all instances stay within a secured VPC with no internet access, and data communication traffic must stay within the AWS network.
How should the Data Science team configure the notebook instance placement to meet these requirements?

A. Associate the Amazon SageMaker notebook with a private subnet in a VP
B. Place the Amazon SageMaker endpoint and S3 buckets within the same VPC.
C. Associate the Amazon SageMaker notebook with a private subnet in a VP
D. Use 1AM policies to grant access to Amazon S3 and Amazon SageMaker.
E. Associate the Amazon SageMaker notebook with a private subnet in a VP
F. Ensure the VPC has S3 VPC endpoints and Amazon SageMaker VPC endpoints attached to it.
G. Associate the Amazon SageMaker notebook with a private subnet in a VP
H. Ensure the VPC has a NAT gateway and an associated security group allowing only outbound connections to Amazon S3 and Amazon SageMaker

Answer: D

NEW QUESTION 103
A Machine Learning Specialist uploads a dataset to an Amazon S3 bucket protected with server-side encryption using AWS KMS.
How should the ML Specialist define the Amazon SageMaker notebook instance so it can read the same dataset from Amazon S3?

A. Define security group(s) to allow all HTTP inbound/outbound traffic and assign those security group(s) tothe Amazon SageMaker notebook instance.
B. onfigure the Amazon SageMaker notebook instance to have access to the VP
C. Grant permission in the KMS key policy to the notebook's KMS role.
D. Assign an IAM role to the Amazon SageMaker notebook with S3 read access to the datase
E. Grant permission in the KMS key policy to that role.
F. Assign the same KMS key used to encrypt data in Amazon S3 to the Amazon SageMaker notebook instance.

Answer: D

NEW QUESTION 104
A company wants to create a data repository in the AWS Cloud for machine learning (ML) projects. The company wants to use AWS to perform complete ML lifecycles and wants to use Amazon S3 for the data storage. All of the company's data currently resides on premises and is 40 in size.
The company wants a solution that can transfer and automatically update data between the on-premises object storage and Amazon S3. The solution must support encryption, scheduling, monitoring, and data integrity validation.
Which solution meets these requirements?

A. Use the S3 sync command to compare the source S3 bucket and the destination S3 bucke
B. Determine which source files do not exist in the destination S3 bucket and which source files were modified.
C. Use AWS Transfer for FTPS to transfer the files from the on-premises storage to Amazon S3.
D. Use AWS DataSync to make an initial copy of the entire datase
E. Schedule subsequent incremental transfers of changing data until the final cutover from on premises to AWS.
F. Use S3 Batch Operations to pull data periodically from the on-premises storag
G. Enable S3 Versioning on the S3 bucket to protect against accidental overwrites.

Answer: C

Explanation:
Configure DataSync to make an initial copy of your entire dataset, and schedule subsequent incremental transfers of changing data until the final cut-over from on-premises to AWS.

NEW QUESTION 108
A retail chain has been ingesting purchasing records from its network of 20,000 stores to Amazon S3 using Amazon Kinesis Data Firehose To support training an improved machine learning model, training records will require new but simple transformations, and some attributes will be combined The model needs lo be retrained daily
Given the large number of stores and the legacy data ingestion, which change will require the LEAST amount of development effort?

A. Require that the stores to switch to capturing their data locally on AWS Storage Gateway for loading into Amazon S3 then use AWS Glue to do the transformation
B. Deploy an Amazon EMR cluster running Apache Spark with the transformation logic, and have the cluster run each day on the accumulating records in Amazon S3, outputting new/transformed records to Amazon S3
C. Spin up a fleet of Amazon EC2 instances with the transformation logic, have them transform the data records accumulating on Amazon S3, and output the transformed records to Amazon S3.
D. Insert an Amazon Kinesis Data Analytics stream downstream of the Kinesis Data Firehouse stream that transforms raw record attributes into simple transformed values using SQL.

Answer: D

NEW QUESTION 110
A data scientist is working on a public sector project for an urban traffic system. While studying the traffic patterns, it is clear to the data scientist that the traffic behavior at each light is correlated, subject to a small stochastic error term. The data scientist must model the traffic behavior to analyze the traffic patterns and reduce congestion.
How will the data scientist MOST effectively model the problem?

A. The data scientist should obtain a correlated equilibrium policy by formulating this problem as a multi-agent reinforcement learning problem.
B. The data scientist should obtain the optimal equilibrium policy by formulating this problem as a single-agent reinforcement learning problem.
C. Rather than finding an equilibrium policy, the data scientist should obtain accurate predictors of traffic flow by using historical data through a supervised learning approach.
D. Rather than finding an equilibrium policy, the data scientist should obtain accurate predictors of traffic flow by using unlabeled simulated data representing the new traffic patterns in the city and applying an unsupervised learning approach.

Answer: D

NEW QUESTION 113
A Machine Learning Specialist has created a deep learning neural network model that performs well on the training data but performs poorly on the test data.
Which of the following methods should the Specialist consider using to correct this? (Select THREE.)

A. Decrease regularization.
B. Increase regularization.
C. Increase dropout.
D. Decrease dropout.
E. Increase feature combinations.
F. Decrease feature combinations.

Answer: BCD

NEW QUESTION 116
A manufacturing company asks its Machine Learning Specialist to develop a model that classifies defective parts into one of eight defect types. The company has provided roughly 100000 images per defect type for training During the injial training of the image classification model the Specialist notices that the validation accuracy is 80%, while the training accuracy is 90% It is known that human-level performance for this type of image classification is around 90%
What should the Specialist consider to fix this issue1?

A. A longer training time
B. Making the network larger
C. Using a different optimizer
D. Using some form of regularization

Answer: D

NEW QUESTION 117
A Machine Learning Specialist previously trained a logistic regression model using scikit-learn on a local machine, and the Specialist now wants to deploy it to production for inference only.
What steps should be taken to ensure Amazon SageMaker can host a model that was trained locally?

A. Build the Docker image with the inference cod

B. Tag the Docker image with the registry hostname and upload it to Amazon ECR.
C. Serialize the trained model so the format is compressed for deploymen
D. Tag the Docker image with the registry hostname and upload it to Amazon S3.
E. Serialize the trained model so the format is compressed for deploymen
F. Build the image and upload it to Docker Hub.
G. Build the Docker image with the inference cod
H. Configure Docker Hub and upload the image to AmazonECR.

Answer: D

NEW QUESTION 120
A machine learning specialist is running an Amazon SageMaker endpoint using the built-in object detection algorithm on a P3 instance for real-time predictions in a company's production application. When evaluating the model's resource utilization, the specialist notices that the model is using only a fraction of the GPU.
Which architecture changes would ensure that provisioned resources are being utilized effectively?

A. Redeploy the model as a batch transform job on an M5 instance.
B. Redeploy the model on an M5 instanc
C. Attach Amazon Elastic Inference to the instance.
D. Redeploy the model on a P3dn instance.
E. Deploy the model onto an Amazon Elastic Container Service (Amazon ECS) cluster using a P3 instance.

Answer: B

Explanation:
https://aws.amazon.com/machine-learning/elastic-inference/

NEW QUESTION 124
A company that runs an online library is implementing a chatbot using Amazon Lex to provide book recommendations based on category. This intent is fulfilled by an AWS Lambda function that queries an Amazon DynamoDB table for a list of book titles, given a particular category. For testing, there are only three categories implemented as the custom slot types: "comedy," "adventure," and "documentary."
A machine learning (ML) specialist notices that sometimes the request cannot be fulfilled because Amazon Lex cannot understand the category spoken by users with utterances such as "funny," "fun," and "humor." The ML specialist needs to fix the problem without changing the Lambda code or data in DynamoDB.
How should the ML specialist fix the problem?

A. Add the unrecognized words in the enumeration values list as new values in the slot type.
B. Create a new custom slot type, add the unrecognized words to this slot type as enumeration values, and use this slot type for the slot.
C. Use the AMAZON.SearchQuery built-in slot types for custom searches in the database.
D. Add the unrecognized words as synonyms in the custom slot type.

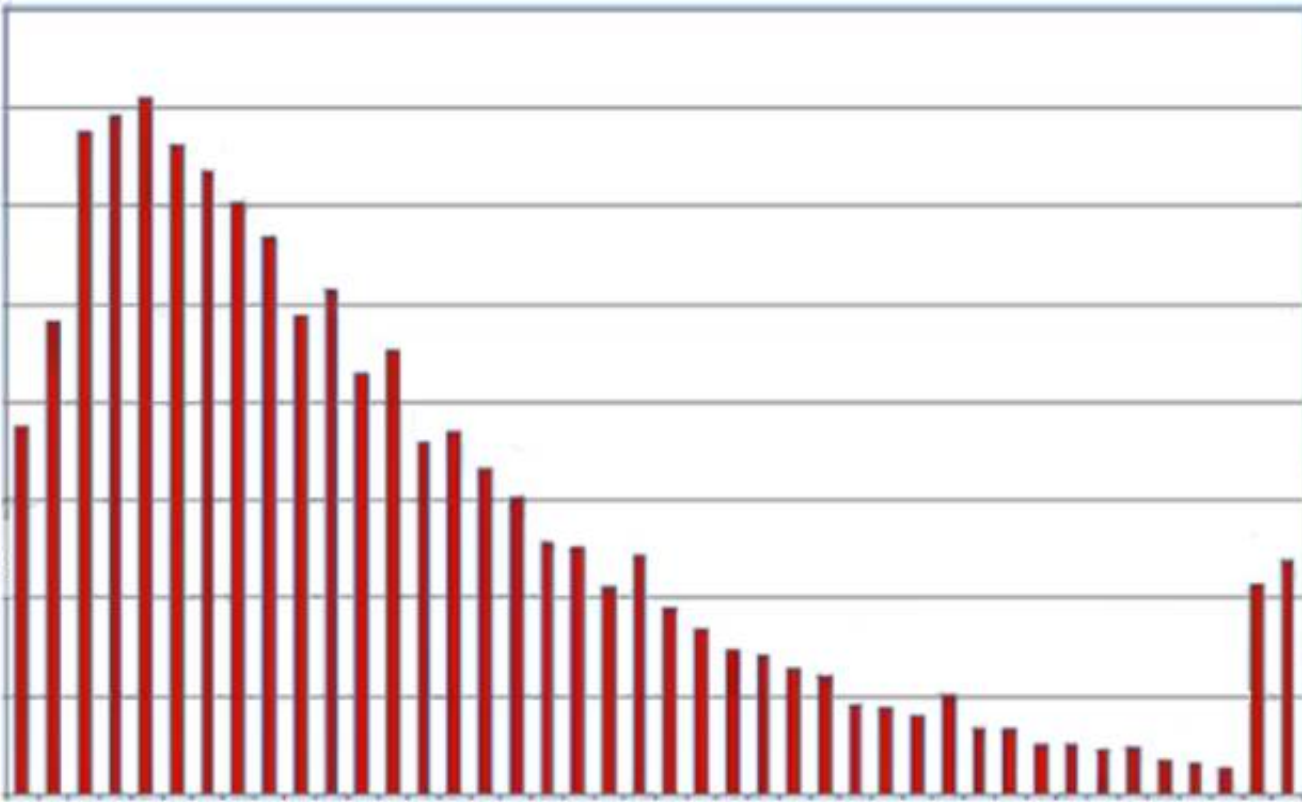**Answer:** C

NEW QUESTION 129
A Machine Learning Specialist is using an Amazon SageMaker notebook instance in a private subnet of a corporate VPC. The ML Specialist has important data stored on the Amazon SageMaker notebook instance's Amazon EBS volume, and needs to take a snapshot of that EBS volume. However the ML Specialist cannot find the Amazon SageMaker notebook instance's EBS volume or Amazon EC2 instance within the VPC.
Why is the ML Specialist not seeing the instance visible in the VPC?

A. Amazon SageMaker notebook instances are based on the EC2 instances within the customer account, but they run outside of VPCs.
B. Amazon SageMaker notebook instances are based on the Amazon ECS service within customer accounts.
C. Amazon SageMaker notebook instances are based on EC2 instances running within AWS serviceaccounts.
D. Amazon SageMaker notebook instances are based on AWS ECS instances running within AWS service accounts.

Answer: C

NEW QUESTION 133
A Data Scientist is building a linear regression model and will use resulting p-values to evaluate the statistical significance of each coefficient. Upon inspection of the dataset, the Data Scientist discovers that most of the features are normally distributed. The plot of one feature in the dataset is shown in the graphic.

What transformation should the Data Scientist apply to satisfy the statistical assumptions of the linear regression model?

A. Exponential transformation
B. Logarithmic transformation
C. Polynomial transformation
D. Sinusoidal transformation

Answer: A

NEW QUESTION 134
A Mobile Network Operator is building an analytics platform to analyze and optimize a company's operations using Amazon Athena and Amazon S3
The source systems send data in CSV format in real lime The Data Engineering team wants to transform the data to the Apache Parquet format before storing it on Amazon S3
Which solution takes the LEAST effort to implement?

A. Ingest .CSV data using Apache Kafka Streams on Amazon EC2 instances and use Kafka Connect S3 toserialize data as Parquet
B. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Glue to convert data into Parquet.
C. Ingest .CSV data using Apache Spark Structured Streaming in an Amazon EMR cluster and use Apache Spark to convert data into Parquet.
D. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Kinesis Data Firehose to convert data into Parquet.

Answer: B

Explanation:

https://medium.com/searce/convert-csv-json-files-to-apache-parquet-using-aws-glue-a760d177b45f https://github.com/ecloudvalley/Building-a-Data-Lake-with-AWS-Glue-and-Amazon-S3

NEW QUESTION 136
A retail company is using Amazon Personalize to provide personalized product recommendations for its customers during a marketing campaign. The company sees a significant increase in sales of recommended items to existing customers immediately after deploying a new solution version, but these sales decrease a short time after deployment. Only historical data from before the marketing campaign is available for training.
How should a data scientist adjust the solution?

A. Use the event tracker in Amazon Personalize to include real-time user interactions.
B. Add user metadata and use the HRNN-Metadata recipe in Amazon Personalize.
C. Implement a new solution using the built-in factorization machines (FM) algorithm in Amazon SageMaker.
D. Add event type and event value fields to the interactions dataset in Amazon Personalize.

Answer: A

NEW QUESTION 141
A Machine Learning Specialist works for a credit card processing company and needs to predict which transactions may be fraudulent in near-real time.
Specifically, the Specialist must train a model that returns the probability that a given transaction may fraudulent.
How should the Specialist frame this business problem?

A. Streaming classification
B. Binary classification
C. Multi-category classification
D. Regression classification

Answer: C

NEW QUESTION 145
A Machine Learning Specialist working for an online fashion company wants to build a data ingestion solution for the company's Amazon S3-based data lake.
The Specialist wants to create a set of ingestion mechanisms that will enable future capabilities comprised of:
• Real-time analytics

• Interactive analytics of historical data
• Clickstream analytics
• Product recommendations
Which services should the Specialist use?

A. AWS Glue as the data dialog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for real-time data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
B. Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for near-realtime data insights; Amazon Kinesis Data Firehose for clickstream analytics; AWS Glue to generate personalized product recommendations
C. AWS Glue as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
D. Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon DynamoDB streams for clickstream analytics; AWS Glue to generate personalized product recommendations

Answer: A

NEW QUESTION 148
A Data Scientist is building a model to predict customer churn using a dataset of 100 continuous numerical features. The Marketing team has not provided any insight about which features are relevant for churn prediction. The Marketing team wants to interpret the model and see the direct impact of relevant features on the model outcome. While training a logistic regression model, the Data Scientist observes that there is a wide gap between the training and validation set accuracy.
Which methods can the Data Scientist use to improve the model performance and satisfy the Marketing team's needs? (Choose two.)

A. Add L1 regularization to the classifier
B. Add features to the dataset
C. Perform recursive feature elimination
D. Perform t-distributed stochastic neighbor embedding (t-SNE)
E. Perform linear discriminant analysis

Answer: BE

NEW QUESTION 153
An agricultural company is interested in using machine learning to detect specific types of weeds in a 100-acre grassland field. Currently, the company uses tractor-mounted cameras to capture multiple images of the field as 10 × 10 grids. The company also has a large training dataset that consists of annotated images of popular weed classes like broadleaf and non-broadleaf docks.
The company wants to build a weed detection model that will detect specific types of weeds and the location of each type within the field. Once the model is ready, it will be hosted on Amazon SageMaker endpoints. The model will perform real-time inferencing using the images captured by the cameras. Which approach should a Machine Learning Specialist take to obtain accurate predictions?

A. Prepare the images in RecordIO format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an image classification algorithm to categorize images into various weed classes.
B. Prepare the images in Apache Parquet format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an object-detection single-shot multibox detector (SSD) algorithm.
C. Prepare the images in RecordIO format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an object-detection single-shot multibox detector (SSD) algorithm.
D. Prepare the images in Apache Parquet format and upload them to Amazon S3. Use Amazon SageMaker to train, test, and validate the model using an image classification algorithm to categorize images into various weed classes.

Answer: C

NEW QUESTION 158
A large consumer goods manufacturer has the following products on sale
• 34 different toothpaste variants
• 48 different toothbrush variants
• 43 different mouthwash variants
The entire sales history of all these products is available in Amazon S3 Currently, the company is using custom-built autoregressive integrated moving average (ARIMA) models to forecast demand for these products The company wants to predict the demand for a new product that will soon be launched
Which solution should a Machine Learning Specialist apply?

A. Train a custom ARIMA model to forecast demand for the new product.
B. Train an Amazon SageMaker DeepAR algorithm to forecast demand for the new product
C. Train an Amazon SageMaker k-means clustering algorithm to forecast demand for the new product.
D. Train a custom XGBoost model to forecast demand for the new product

Answer: B

Explanation:
The Amazon SageMaker DeepAR forecasting algorithm is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNN). Classical forecasting methods, such as autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS), fit a single model to each individual time series. They then use that model to extrapolate the time series into the future.

NEW QUESTION 161
A Machine Learning Specialist receives customer data for an online shopping website. The data includes demographics, past visits, and locality information. The Specialist must develop a machine learning approach to identify the customer shopping patterns, preferences and trends to enhance the website for better service and smart recommendations.
Which solution should the Specialist recommend?

A. Latent Dirichlet Allocation (LDA) for the given collection of discrete data to identify patterns in thecustomer database.
B. A neural network with a minimum of three layers and random initial weights to identify patterns in the customer database

C. Collaborative filtering based on user interactions and correlations to identify patterns in the customer database
D. Random Cut Forest (RCF) over random subsamples to identify patterns in the customer database

Answer: C

NEW QUESTION 166
A telecommunications company is developing a mobile app for its customers. The company is using an Amazon SageMaker hosted endpoint for machine learning model inferences.
Developers want to introduce a new version of the model for a limited number of users who subscribed to a preview feature of the app. After the new version of the model is tested as a preview, developers will evaluate its accuracy. If a new version of the model has better accuracy, developers need to be able to gradually release the new version for all users over a fixed period of time.
How can the company implement the testing model with the LEAST amount of operational overhead?

A. Update the ProductionVariant data type with the new version of the model by using the CreateEndpointConfig operation with the InitialVariantWeight parameter set to 0. Specify the TargetVariant parameter for InvokeEndpoint calls for users who subscribed to the preview featur
B. When the new version of the model is ready for release, gradually increase InitialVariantWeight until all users have the updated version.
C. Configure two SageMaker hosted endpoints that serve the different versions of the mode
D. Create an Application Load Balancer (ALB) to route traffic to both endpoints based on the TargetVariant query string paramete
E. Reconfigure the app to send the TargetVariant query string parameter for users who subscribed to the preview featur
F. When the new version of the model is ready for release, change the ALB's routing algorithm to weighted until all users have the updated version.
G. Update the DesiredWeightsAndCapacity data type with the new version of the model by using the UpdateEndpointWeightsAndCapacities operation with the DesiredWeight parameter set to 0. Specify the TargetVariant parameter for InvokeEndpoint calls for users who subscribed to the preview featur
H. When the new version of the model is ready for release, gradually increase DesiredWeight until all users have the updated version.
I. Configure two SageMaker hosted endpoints that serve the different versions of the mode
J. Create an Amazon Route 53 record that is configured with a simple routing policy and that points to the current version of the mode
K. Configure the mobile app to use the endpoint URL for users who subscribed to the preview feature and to use the Route 53 record for other user
L. When the new version of the model is ready for release, add a new model version endpoint to Route 53, and switch the policy to weighted until all users have the updated version.

Answer: D

NEW QUESTION 170
An employee found a video clip with audio on a company's social media feed. The language used in the video is Spanish. English is the employee's first language, and they do not understand Spanish. The employee wants to do a sentiment analysis.
What combination of services is the MOST efficient to accomplish the task?

A. Amazon Transcribe, Amazon Translate, and Amazon Comprehend
B. Amazon Transcribe, Amazon Comprehend, and Amazon SageMaker seq2seq
C. Amazon Transcribe, Amazon Translate, and Amazon SageMaker Neural Topic Model (NTM)
D. Amazon Transcribe, Amazon Translate, and Amazon SageMaker BlazingText

Answer: A

NEW QUESTION 171
A Machine Learning Specialist is working with a large cybersecurily company that manages security events in real time for companies around the world The cybersecurity company wants to design a solution that will allow it to use machine learning to score malicious events as anomalies on the data as it is being ingested The company also wants be able to save the results in its data lake for later processing and analysis
What is the MOST efficient way to accomplish these tasks'?

A. Ingest the data using Amazon Kinesis Data Firehose, and use Amazon Kinesis Data Analytics Random Cut Forest (RCF) for anomaly detection Then use Kinesis Data Firehose to stream the results to Amazon S3
B. Ingest the data into Apache Spark Streaming using Amazon EM
C. and use Spark MLlib with k-means to perform anomaly detection Then store the results in an Apache Hadoop Distributed File System (HDFS) using Amazon EMR with a replication factor of three as the data lake
D. Ingest the data and store it in Amazon S3 Use AWS Batch along with the AWS Deep Learning AMIs to train a k-means model using TensorFlow on the data in Amazon S3.
E. Ingest the data and store it in Amazon S3. Have an AWS Glue job that is triggered on demand transform the new data Then use the built-in Random Cut Forest (RCF) model within Amazon SageMaker to detect anomalies in the data

Answer: A

NEW QUESTION 176
Which of the following metrics should a Machine Learning Specialist generally use to compare/evaluate machine learning classification models against each other?

A. Recall
B. Misclassification rate
C. Mean absolute percentage error (MAPE)
D. Area Under the ROC Curve (AUC)

Answer: D

NEW QUESTION 177
A Machine Learning Specialist is assigned a TensorFlow project using Amazon SageMaker for training, and needs to continue working for an extended period with no Wi-Fi access.
Which approach should the Specialist use to continue working?

A. Install Python 3 and boto3 on their laptop and continue the code development using that environment.
B. Download the TensorFlow Docker container used in Amazon SageMaker from GitHub to their local environment, and use the Amazon SageMaker Python SDK

to test the code.
C. Download TensorFlow from tensorflow.org to emulate the TensorFlow kernel in the SageMaker environment.
D. Download the SageMaker notebook to their local environment then install Jupyter Notebooks on their laptop and continue the development in a local notebook.

Answer: D

NEW QUESTION 179
A company's Machine Learning Specialist needs to improve the training speed of a time-series forecasting model using TensorFlow. The training is currently implemented on a single-GPU machine and takes approximately 23 hours to complete. The training needs to be run daily.
The model accuracy js acceptable, but the company anticipates a continuous increase in the size of the training data and a need to update the model on an hourly, rather than a daily, basis. The company also wants to minimize coding effort and infrastructure changes
What should the Machine Learning Specialist do to the training solution to allow it to scale for future demand?

A. Do not change the TensorFlow cod
B. Change the machine to one with a more powerful GPU to speed up the training.
C. Change the TensorFlow code to implement a Horovod distributed framework supported by Amazon SageMake
D. Parallelize the training to as many machines as needed to achieve the business goals.
E. Switch to using a built-in AWS SageMaker DeepAR mode
F. Parallelize the training to as many machines as needed to achieve the business goals.
G. Move the training to Amazon EMR and distribute the workload to as many machines as needed to achieve the business goals.

Answer: B

NEW QUESTION 182
A Machine Learning Specialist is working for an online retailer that wants to run analytics on every customer visit, processed through a machine learning pipeline. The data needs to be ingested by Amazon Kinesis Data Streams at up to 100 transactions per second, and the JSON data blob is 100 KB in size.
What is the MINIMUM number of shards in Kinesis Data Streams the Specialist should use to successfully ingest this data?

A. 1 shards
B. 10 shards
C. 100 shards
D. 1,000 shards

Answer: B

NEW QUESTION 183
A Machine Learning Specialist is applying a linear least squares regression model to a dataset with 1 000 records and 50 features Prior to training, the ML Specialist notices that two features are perfectly linearly dependent
Why could this be an issue for the linear least squares regression model?

A. It could cause the backpropagation algorithm to fail during training
B. It could create a singular matrix during optimization which fails to define a unique solution
C. It could modify the loss function during optimization causing it to fail during training
D. It could introduce non-linear dependencies within the data which could invalidate the linear assumptions of the model

Answer: C

NEW QUESTION 186
A Machine Learning Specialist wants to determine the appropriate SageMakerVariant Invocations Per Instance setting for an endpoint automatic scaling configuration. The Specialist has performed a load test on a single instance and determined that peak requests per second (RPS) without service degradation is about 20 RPS As this is the first deployment, the Specialist intends to set the invocation safety factor to 0 5
Based on the stated parameters and given that the invocations per instance setting is measured on a per-minute basis, what should the Specialist set as the sageMakervariantinvocationsPerinstance setting?

A. 10
B. 30
C. 600
D. 2,400

Answer: C

NEW QUESTION 191
A machine learning specialist needs to analyze comments on a news website with users across the globe. The specialist must find the most discussed topics in the comments that are in either English or Spanish.
What steps could be used to accomplish this task? (Choose two.)

A. Use an Amazon SageMaker BlazingText algorithm to find the topics independently from language.Proceed with the analysis.
B. Use an Amazon SageMaker seq2seq algorithm to translate from Spanish to English, if necessar
C. Use aSageMaker Latent Dirichlet Allocation (LDA) algorithm to find the topics.
D. Use Amazon Translate to translate from Spanish to English, if necessar
E. Use Amazon Comprehend topic modeling to find the topics.
F. Use Amazon Translate to translate from Spanish to English, if necessar
G. Use Amazon Lex to extract topics form the content.
H. Use Amazon Translate to translate from Spanish to English, if necessar
I. Use Amazon SageMaker Neural Topic Model (NTM) to find the topics.

Answer: B

NEW QUESTION 192
A Machine Learning Specialist is building a model to predict future employment rates based on a wide range of economic factors While exploring the data, the Specialist notices that the magnitude of the input features vary greatly The Specialist does not want variables with a larger magnitude to dominate the model
What should the Specialist do to prepare the data for model training'?

A. Apply quantile binning to group the data into categorical bins to keep any relationships in the data by replacing the magnitude with distribution
B. Apply the Cartesian product transformation to create new combinations of fields that are independent of the magnitude
C. Apply normalization to ensure each field will have a mean of 0 and a variance of 1 to remove any significant magnitude
D. Apply the orthogonal sparse Diagram (OSB) transformation to apply a fixed-size sliding window to generate new features of a similar magnitude.

Answer: C

NEW QUESTION 195
A retail company is selling products through a global online marketplace. The company wants to use machine learning (ML) to analyze customer feedback and identify specific areas for improvement. A developer has built a tool that collects customer reviews from the online marketplace and stores them in an Amazon S3 bucket. This process yields a dataset of 40 reviews. A data scientist building the ML models must identify additional sources of data to increase the size of the dataset.
Which data sources should the data scientist use to augment the dataset of reviews? (Choose three.)

A. Emails exchanged by customers and the company's customer service agents
B. Social media posts containing the name of the company or its products
C. A publicly available collection of news articles
D. A publicly available collection of customer reviews
E. Product sales revenue figures for the company
F. Instruction manuals for the company's products

Answer: BDF

NEW QUESTION 199
During mini-batch training of a neural network for a classification problem, a Data Scientist notices that training accuracy oscillates What is the MOST likely cause of this issue?

A. The class distribution in the dataset is imbalanced
B. Dataset shuffling is disabled
C. The batch size is too big
D. The learning rate is very high

Answer: B

NEW QUESTION 202
A data scientist needs to identify fraudulent user accounts for a company's ecommerce platform. The company wants the ability to determine if a newly created account is associated with a previously known fraudulent user. The data scientist is using AWS Glue to cleanse the company's application logs during ingestion.
Which strategy will allow the data scientist to identify fraudulent accounts?

A. Execute the built-in FindDuplicates Amazon Athena query.
B. Create a FindMatches machine learning transform in AWS Glue.
C. Create an AWS Glue crawler to infer duplicate accounts in the source data.
D. Search for duplicate accounts in the AWS Glue Data Catalog.

Answer: B

NEW QUESTION 207
A Machine Learning Specialist has built a model using Amazon SageMaker built-in algorithms and is not getting expected accurate results The Specialist wants to use hyperparameter optimization to increase the model's accuracy
Which method is the MOST repeatable and requires the LEAST amount of effort to achieve this?

A. Launch multiple training jobs in parallel with different hyperparameters
B. Create an AWS Step Functions workflow that monitors the accuracy in Amazon CloudWatch Logs and relaunches the training job with a defined list of hyperparameters
C. Create a hyperparameter tuning job and set the accuracy as an objective metric.
D. Create a random walk in the parameter space to iterate through a range of values that should be used for each individual hyperparameter

Answer: B

NEW QUESTION 208
A Data Science team is designing a dataset repository where it will store a large amount of training data commonly used in its machine learning models. As Data Scientists may create an arbitrary number of new datasets every day the solution has to scale automatically and be cost-effective. Also, it must be possible to explore the data using SQL.
Which storage scheme is MOST adapted to this scenario?

A. Store datasets as files in Amazon S3.
B. Store datasets as files in an Amazon EBS volume attached to an Amazon EC2 instance.
C. Store datasets as tables in a multi-node Amazon Redshift cluster.
D. Store datasets as global tables in Amazon DynamoDB.

Answer: A

NEW QUESTION 209

A manufacturer is operating a large number of factories with a complex supply chain relationship where unexpected downtime of a machine can cause production to stop at several factories. A data scientist wants to analyze sensor data from the factories to identify equipment in need of preemptive maintenance and then dispatch a service team to prevent unplanned downtime. The sensor readings from a single machine can include up to 200 data points including temperatures, voltages, vibrations, RPMs, and pressure readings.

To collect this sensor data, the manufacturer deployed Wi-Fi and LANs across the factories. Even though many factory locations do not have reliable or high-speed internet connectivity, the manufacturer would like to maintain near-real-time inference capabilities.

Which deployment architecture for the model will address these business requirements?

A. Deploy the model in Amazon SageMake
B. Run sensor data through this model to predict which machinesneed maintenance.
C. Deploy the model on AWS IoT Greengrass in each factor
D. Run sensor data through this model to infer which machines need maintenance.
E. Deploy the model to an Amazon SageMaker batch transformation jo
F. Generate inferences in a daily batch report to identify machines that need maintenance.
G. Deploy the model in Amazon SageMaker and use an IoT rule to write data to an Amazon DynamoDB table.Consume a DynamoDB stream from the table with an AWS Lambda function to invoke the endpoint.

Answer: B

Explanation:
https://aws.amazon.com/blogs/iot/industrial-iot-from-condition-based-monitoring-to-predictive-quality-to-digitiz https://aws.amazon.com/blogs/iot/using-aws-iot-for-predictive-maintenance/

NEW QUESTION 214

A data scientist is using an Amazon SageMaker notebook instance and needs to securely access data stored in a specific Amazon S3 bucket.
How should the data scientist accomplish this?

A. Add an S3 bucket policy allowing GetObject, PutObject, and ListBucket permissions to the AmazonSageMaker notebook ARN as principal.
B. Encrypt the objects in the S3 bucket with a custom AWS Key Management Service (AWS KMS) key that only the notebook owner has access to.
C. Attach the policy to the IAM role associated with the notebook that allows GetObject, PutObject, and ListBucket operations to the specific S3 bucket.
D. Use a script in a lifecycle configuration to configure the AWS CLI on the instance with an access key ID and secret.

Answer: C

NEW QUESTION 217

For the given confusion matrix, what is the recall and precision of the model?

A. Recall = 0.92 Precision = 0.84
B. Recall = 0.84 Precision = 0.8
C. Recall = 0.92 Precision = 0.8
D. Recall = 0.8 Precision = 0.92

Answer: C

NEW QUESTION 218
......

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questons and Answers in PDF Format

AWS-Certified-Machine-Learning-Specialty Practice Exam Features:

* AWS-Certified-Machine-Learning-Specialty Questions and Answers Updated Frequently

* AWS-Certified-Machine-Learning-Specialty Practice Questions Verified by Expert Senior Certified Staff

* AWS-Certified-Machine-Learning-Specialty Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* AWS-Certified-Machine-Learning-Specialty Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
Order The AWS-Certified-Machine-Learning-Specialty Practice Test Here